


```
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 230
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-893-737-230

Query Match      100.0%; Score 770; DB 3; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.7e-77;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVFTLLVLLVLLLTALCKALSQSLPYTLRPOSSLSFLITDIKKIDIOYFLPTGG 60
   |||||
Db 1 MVVFTLLVLLVLLLTALCKALSQSLPYTLRPOSSLSFLITDIKKIDIOYFLPTGG 60
   |||||

QY 61 KCLHLRLTGORAFCLLEFLPWCNGIIESLVGLNERKVLGGSSQOMGEARRGMWEVFP 120
   |||||
Db 61 KCLHLRLTGORAFCLLEFLPWCNGIIESLVGLNERKVLGGSSQOMGEARRGMWEVFP 120
   |||||

QY 121 LEIGRPEAGALQRLPQNSALLACRCAGAY 150
   |||||
Db 121 LEIGRPEAGALQRLPQNSALLACRCAGAY 150
   |||||

RESULT 3
US-10-777-578-2
; Sequence 2, Application US/10777578
; Publication No. US20040147722A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: NOVEL SECRETED PROTEINS
; FILE REFERENCE: 01-26
; CURRENT APPLICATION NUMBER: US/10/777,578
; PRIOR FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: US/09/895,836
; PRIOR FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-777-578-2

Query Match      100.0%; Score 770; DB 4; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.7e-77;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVFTLLVLLVLLLTALCKALSQSLPYTLRPOSSLSFLITDIKKIDIOYFLPTGG 60
   |||||
Db 1 MVVFTLLVLLVLLLTALCKALSQSLPYTLRPOSSLSFLITDIKKIDIOYFLPTGG 60
   |||||

QY 61 KCLHLRLTGORAFCLLEFLPWCNGIIESLVGLNERKVLGGSSQOMGEARRGMWEVFP 120
   |||||
Db 61 KCLHLRLTGORAFCLLEFLPWCNGIIESLVGLNERKVLGGSSQOMGEARRGMWEVFP 120
   |||||

QY 121 LEIGRPEAGALQRLPQNSALLACRCAGAY 150
   |||||
Db 121 LEIGRPEAGALQRLPQNSALLACRCAGAY 150
   |||||

RESULT 4
US-10-970-713-230
; Sequence 230, Application US/10970713
; Publication No. US20050214791A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
; FILE REFERENCE: 00-41
; CURRENT APPLICATION NUMBER: US/10/970,713
; CURRENT FILING DATE: 2004-10-20
; PRIOR APPLICATION NUMBER: US/09/893,737
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 329
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 230
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-970-713-230

Query Match      100.0%; Score 770; DB 5; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.7e-77;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVFTLLVLLVLLLTALCKALSQSLPYTLRPOSSLSFLITDIKKIDIOYFLPTGG 60
   |||||
Db 1 MVVFTLLVLLVLLLTALCKALSQSLPYTLRPOSSLSFLITDIKKIDIOYFLPTGG 60
   |||||

QY 61 KCLHLRLTGORAFCLLEFLPWCNGIIESLVGLNERKVLGGSSQOMGEARRGMWEVFP 120
   |||||
Db 61 KCLHLRLTGORAFCLLEFLPWCNGIIESLVGLNERKVLGGSSQOMGEARRGMWEVFP 120
   |||||

QY 121 LEIGRPEAGALQRLPQNSALLACRCAGAY 150
   |||||
Db 121 LEIGRPEAGALQRLPQNSALLACRCAGAY 150
   |||||

RESULT 5
US-10-450-763-59832
; Sequence 59832, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 59832
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(246)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-59832

Query Match      21.4%; Score 164.5; DB 5; Length 246;
Best Local Similarity 60.3%; Pred. No. 1.3e-09;
Matches 47; Conservative 2; Mismatches 22; Indels 7; Gaps 4;

QY 64 HLRLTGORAFCLLEFLPWCNGIIESLVGLNERKVLGGSSQOMGEARRGMWEVFP 121
   |||||
Db 58 HSLSQORAFCLLGFLPWHRRIRSHVGLNECKVLLSGRSSQOMGEARRGMWEVFP 117
   |||||
QY 122 ELGRPEAGALQRLPQNS 139
   |||||
```

GenCore version 5.1.9
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2007, 15:01:16 ; Search time 38 Seconds
(without alignments)
379.803 Million cell updates/sec

Title: US-10-777-578-2

Perfect score: 770

Sequence: 1 MWVFTLLVLLLLLTALC.....LQRLPQPNALLACRCAGAY 150

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_80:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83.5	10.8	254	H97081	proline/glycine be
2	75.5	9.8	483	A38560	nitrate transport
3	73	9.5	492	F83462	hypothetical prote
4	73	9.5	1163	F84669	probable chromosom
5	72.5	9.4	326	B46108	outer capsid prote
6	72.5	9.4	343	D75260	probable carboxydr
7	71.5	9.3	253	G30275	hypothetical prote
8	71	9.2	406	S76451	hypothetical prote
9	70.5	9.2	326	1 VGXRB	glycoprotein VP7 p
10	70.5	9.2	326	A44895	major glycoprotein
11	70	9.1	442	T06476	ppf-1 protein - ga
12	69.5	9.0	474	S75568	isochorismate synt
13	69.5	9.0	649	H64476	lipocate protein li
14	69	9.0	216	S70786	probable transcrip
15	69	9.0	253	A90509	cobalamin synthase
16	69	9.0	1212	T00332	hypothetical prote
17	68.5	8.9	532	T44788	phospho-prenol glu
18	68.5	8.9	597	E40201	artifact-warning s
19	68.5	8.9	1627	A82109	two-component hybr
20	68	8.8	309	AC1349	oligopeptide ABC t
21	68	8.8	618	T48193	hypothetical prote
22	68	8.8	1064	B86465	probable protein k
23	68	8.8	1932	T25525	hypothetical prote
24	67.5	8.8	260	C69893	probable enoyl-CoA
25	67.5	8.8	326	1 VGXR49	glycoprotein VP7 p
26	67.5	8.8	326	1 VGXR7H	glycoprotein VP7 p
27	67.5	8.8	326	1 VGXRMD	glycoprotein VP7 p
28	67.5	8.8	326	1 VGXRWA	glycoprotein VP7 p
29	67.5	8.8	326	2 S25546	outer capsid prote

30	67.5	8.8	1121	2 C82120	transcription regu
31	67	8.7	304	2 T52079	probable zinc fing
32	67	8.7	386	2 AD2649	ABC transporter, m
33	67	8.7	386	2 C97431	alpha-glucosides t
34	67	8.7	409	2 T03718	suppressor 2 prote
35	67	8.7	499	1 S66677	thioredoxin-disulf
36	67	8.7	551	2 D71969	1-lactate permease
37	67	8.7	1354	2 T48198	hypothetical prote
38	66.5	8.6	254	1 G64045	3-deoxy-manno-octu
39	66.5	8.6	302	2 T31535	hypothetical prote
40	66.5	8.6	326	1 JQ1442	glycoprotein VP7 p
41	66.5	8.6	326	1 JQ1443	glycoprotein VP7 p
42	66.5	8.6	326	1 VGXRAB	glycoprotein VP7 p
43	66.5	8.6	326	1 VGXR2S	glycoprotein VP7 p
44	66.5	8.6	326	1 VGXR3S	glycoprotein VP7 p
45	66.5	8.6	326	1 VGXRDS	glycoprotein VP7 p

ALIGNMENTS

RESULT 1

H97081

proline/glycine betaine ABC transport system, ATPase component CAC1475 [imported] - C1
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: H97081

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Le
J. Bacteriol. 183, 4823-4838, 2001
J. Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: H97081

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-254 <UNK>

A:Cross-references: UNIPROT:Q97J11; UNIPARC:UPI00000CA1FA; GB:AE001437; PIDN:AAK79443.

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC1475

Query Match 10.8%; Score 83.5; DB 2; Length 254;

Best Local Similarity 32.5%; Pred. No. 0.69;

Matches 37; Conservative 10; Mismatches 42; Indels 25; Gaps 6;

Qy 17 TALCKALSQLPYLYRPSQSLFLLITDIIKKIDIQFLPLTG-----GKCLHLRLTQQ 70

Db 45 TLLKMINR-----LYEPDKGSIVLFNEDIKKIDVVKLRRSIGYVIOQVGLFPHMTIANN 99

Qy 71 RAFCLIFLPCWNGIIIES-----LVGLE-NERK-----VLGGSSSQMGEAR 111

Db 100 IA-TVPKLLKWKREIEKRIDELHLVGLFNPFKKYPSPQLSGGQQQRIGLAR 152

RESULT 2

A38560

nitrate transport protein crna - Emericella nidulans

C:Species: Emericella nidulans, Aspergillus nidulans

C:Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 09-Jul-2004

C:Accession: A38560

R:Unkles, S.E.; Hawker, K.L.; Grieve, C.; Campbell, E.I.; Montague, P.; Kinghorn, J.R.

Proc. Natl. Acad. Sci. U.S.A. 88, 204-208, 1991

A:Title: crna encodes a nitrate transporter in Aspergillus nidulans.

A:Reference number: A38560; MUID:91095428; PMID:1986367

A:Accession: A38560

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-483 <UNK>

A:Cross-references: UNIPROT:P22152; UNIPARC:UPI00000178CA8; GB:M57647

C:Superfamily: nitrate transporter component

C:Keywords: transmembrane protein

Query Match 9.8%; Score 75.5; DB 2; Length 483;

A_Map position: 2

Query Match 9.5%; Score 73; DB 2; Length 1163;
Best Local Similarity 25.5%; Pred. No. 38;
Matches 28; Conservative 27; Mismatches 37; Indels 18; Gaps 5;

QY 24 SQSPLVTLYRPOSSLSFLIITDIKKIDIOYFPLPTGGKCLH-----LRLE---TGORAF 73
 |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
DB 131 SPANPYVVVQQGKTASLTLMKDIERLDL--LKKEIGGTRVYEERRRSRLRMQETGNKRK 187
 |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:

QY 74 CILEFLPWCNGIIIESVLGLNERKVLSGGSSQQMGFAARRGMWEVFPLEL 123
 |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
DB 188 QIIIEVHYLD--ERLRELDEEKLRR-KYQLDKQRKSLEYTIYDKEL 232
 |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:

RESULT 5

B46108
outer capsid protein VP7 - human rotavirus (strain I321)
C:Species: human rotavirus
C>Date: 07-Apr-1994 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C/Accession: B46108
R/Das, M.; Dunn, S.J.; Woode, G.N.; Greenberg, H.B.; Rao, C.D.
Virology 194, 374-379, 1993
A>Title: Both surface proteins (VP4 and VP7) of an asymptomatic neonatal rotavirus st1
A:Reference number: A46108; UID:93242771; PMID:8386881
A:Contents: I321, serotype 10
A:Accession: B46108
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-326 <DAS>
A:Cross-references: UNIPROT:O86188; UNIPARC:UIPI00001786EF; GB:L07658; NID:g310326; P11
A>Note: sequence extracted from NCBI backbone (NCBIN:130701, NCPIP:130693)
A:Note: the sequence in GenBank entry ROLVP7B, release 109, (PID:g310325) has the cod
C:Superfamily: rotavirus glycoprotein VP7

Query Match 9.4%; Score 72.5; DB 2; Length 326;
Best Local Similarity 30.0%; Pred. No. 12;
Matches 18; Conservative 14; Mismatches 17; Indels 11; Gaps 2;

QY 5 TLLLVLLLLLTALCKALSPLTYLYRPOSSLFLIITDIKKIDIQ-----YFLPLTG 59
 |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
DB 8 TFLIVLIILFNLIKSTIRMDYYIK-----FLIIVTIASIVNPQNINLPITG 61
 |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:

RESULT 6

D75260
probable carbohydrate kinase - Deinococcus radiodurans (strain RL)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C/Accession: D75260
R/CWhite, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.M.; Smith, H.O.; Yenter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A>Title: Genome sequence of the radiosistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; UID:20036896; PMID:10567266
A/Accession: D75260
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-343 <WHI>
A:Cross-references: UNIPROT:Q9RRD9; UNIPARC:UIPI00000D3PDE; GB:AE002084; GB:AE000513;
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2553
A:Map position: 1
C:Superfamily: ribokinase

Query Match 9.4%; Score 72.5; DB 2; Length 343;
Best Local Similarity 30.5%; Pred. No. 12;
Matches 29; Conservative 12; Mismatches 33; Indels 21; Gaps 3;

QY 65 LRUTGORACILFEFPWCVNGIITESVLGLENERKVLSGGSSQQMGFAARRGMWEVFPLELG 124
 |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:

Db 82 LRAEGVRAE-VLASAAHPTGVILALIDRRQORAMLTG-----QGADWELLPEELP 130

QY 125 RP-----EAGALQRLPQNSALLACRCAGA 149

Db 131 RDLSSAGHLHLTAWSLFRDPRAALEAARIAKA 165

RESULT 7

G90275

Hypothetical protein SSO1215 [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004

C:Accession: G90275

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Sassen, C.W.; Van der Oost, J.
submitted to Genbank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: G90275

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-253 <KUR>

A:Cross-references: UNIPROT:097YU2; UNIPARC:UPI00000643AE; GB:AB006641; NID:gi13814412; E

C:Genetics:

A:Gene: SSO1215

Query Match 9.3%; Score 71.5; DB 2; Length 253;
Best Local Similarity 32.1%; Pred. No. 11;
Matches 26; Conservative 14; Mismatches 34; Indels 7; Gaps 3;

QY 2 VVFTLLVLLLLLTALCKALSQSLPYLYRQSSLSFLITDIKKIDIQFLPLTGK 61

Db 96 VVFTLGVLLVLSLSYAK-----IPEENFTLKLKIDMEVKRIRI-VSLPLLSK 149

QY 62 CUHLRLTQO-RAPCILEFLPW 81

Db 150 ILWLCRVEKIRELSILKTKDW 170

RESULT 8

S76451

Hypothetical protein - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004

C:Accession: S76451

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.

A:Reference number: S74322; MUID:97061201; PMID:9905231

A:Accession: S76451

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-406 <KAN>

A:Cross-references: UNIPROT:P74479; UNIPARC:UPI000000C1038; EMBL:D90915; GB:AB001139; NID

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Start codon: GTG

Query Match 9.2%; Score 71; DB 2; Length 406;
Best Local Similarity 23.3%; Pred. No. 21;
Matches 31; Conservative 22; Mismatches 34; Indels 46; Gaps 6;

QY 11 LLLLLLTALCKALSQSLPYLYRQSSLSFLII-----TDIKKIDIQF--LPLTGG 60

Db 41 LLLVLAVCLFGATNSMAGWLY-VISGISFALLITAAIIPWLSRLQLRFPQIPVSG 99

QY 61 KCHLRLTQORAFCLIFLPCWNGIIESLVGLNERKVLSSQSQMGEGRMBE-WEVF 119

Db 100 EDLTITLR-----LHNHKK-----EAKNLLQWVDVL 125

Db 120 PLELGRPEAGALQ 132

Db 126 PTGLORPOGEAVE 138

RESULT 9

VGXRB

glycoprotein VP7 precursor - bovine rotavirus A (serotype 10 strain B223)

C:Species: bovine rotavirus A

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999

C:Accession: A38510

R:Xu, L.; Harbourn, D.; McCrae, M.A.

J. Gen. Virol. 72, 177-180, 1991

A:Title: Sequence of the gene encoding the major neutralization antigen (VP7) of serot

A:Reference number: A38510; MUID:91116309; PMID:1703560

A:Accession: A38510

A:Molecule type: genomic RNA

A:Residues: 1-326 <XUL>

A:Cross-references: UNIPARC:UPI00000619B6; EMBL:X52650; NID:G61512; PIDN:CAA36875.1; P

C:Genetics:

A:Map position: segment 9

C:Superfamily: rotavirus glycoprotein VP7

C:Keywords: coat protein; Glycoprotein; transmembrane protein

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-326/Product: glycoprotein VP7 #status predicted <GPV>

F:32-48/Region: hydrophobic #status predicted

F:69,145/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.2%; Score 70.5; DB 1; Length 326;
Best Local Similarity 30.0%; Pred. No. 19;
Matches 18; Conservative 14; Mismatches 17; Indels 11; Gaps 2;

QY 5 TLLVLLVLLLLLTALCKALSQSLPYLYRQSSLSFLITDIKKIDIQ-----YFLPLTG 59

Db 8 TFLYLISIIILLNVLKSLITRMDYIYK-----FLIVTITSIVNAQNYGINLPTG 61

RESULT 10

A44895

major glycoprotein VP7 - bovine rotavirus

C:Species: bovine rotavirus

C>Date: 31-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C:Accession: A44895

R:Huang, J. A.; Nagesha, H. S.; Snodgrass, D. R.; Holmes, I. H.

J. Clin. Microbiol. 30, 85-92, 1992

A:Title: Molecular and serological analyses of two bovine rotaviruses (B-11 and B-60)

A:Reference number: A44895; MUID:92129681; PMID:1310336

A:Accession: A44895

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-326 <HUA>

A:Cross-references: UNIPROT:Q86192; UNIPARC:UPI000000F45B0; GB:M64679; NID:G333829; PID

A:Experimental source: strain B-11

A>Note: sequence extracted from NCBI backbone (NCBI:78093, NCBIP:78154)

C:Superfamily: rotavirus glycoprotein VP7

C:Keywords: glycoprotein

Query Match 9.2%; Score 70.5; DB 2; Length 326;
Best Local Similarity 30.0%; Pred. No. 19;
Matches 18; Conservative 14; Mismatches 17; Indels 11; Gaps 2;

QY 5 TLLVLLVLLLLLTALCKALSQSLPYLYRQSSLSFLITDIKKIDIQ-----YFLPLTG 59

Db 8 TFLYLISIIILLNVLKSLITRMDYIYK-----FLIVTITSIVNAQNYGINLPTG 61

RESULT 11

T06476

ppf-1 protein - garden pea

C:Species: Pisum sativum (garden pea)

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004

C:Accession: T06476

R;Zhu, Y.; Zhang, Y.; Luo, J.; Davies, P.J.; Ho, D.T.H.

Gene 208, 1-6, 1998
A;Title: PPF-1, a post-floral-specific gene expressed in short-day-grown G2 pea, may be
A;Reference number: Z15705; MUID:98147997; PMID:9479033
A;Accession: T06476
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-442 <ZHU>
A;Cross-references: UNIPROT:Q9FY06; UNIPARC:UPI00000ABB7B; EMBL:Y12618; NID:G4218522; P
A;Experimental source: cv. G2
C;Genetics:
A;Gene: ppf-1
F;93-330/Domain: stage III sporulation protein homology <SPOR>

Query Match 9.1%; Score 70; DB 2; Length 442;
Best Local Similarity 25.2%; Pred. No. 29;
Matches 35; Conservative 24; Mismatches 50; Indels 30; Gaps 7;

QY 13 LLLLTALCKALSQSLPYLYRPOSSLSFL-LITDIKKIDIOYFLPLTGGKCLHLRLTGQR 71

Db 122 IILITVIVKA--ATLEPLTKQVESTLANQNLQPKIKAIQERY-----AGNQERIQLETSR 174

QY 72 AF-----CI--LEFLPWCNGIIESLVGLENERKV-----LSGSSQQMGEA 110

Db 175 LYTQAGVNPAGCLPTLATIPVWIGLYQALSNVANEGLLTGFWLPSLGGPTSIARQS 234

QY 111 RRGMEWEVFPLELGRPEAG 129

Db 235 GSGISW-LFPFVDGHPPLG 252

RESULT 12

S75568
Isochorismate synthase (EC 5.4.99.6) - *Synechocystis* sp. (strain PCC 6803)

N;Alternate names: protein slr0817

C;Species: *Synechocystis* sp.

A;Variety: PCC 6803

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

R;Accession: S75568

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

S.

A;Reference number: S74322; MUID:97061201; PMID:8905231

A;Accession: S75568

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-474 <KAN>

A;Cross-references: UNIPARC:UPI00000D3511; EMBL:D90911; GB:AB001339; NID:gl653083; PIDN:

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C;Genetics:

A;Gene: entC

C;Superfamily: isochorismate synthase

C;Keywords: intramolecular transferase; isomerase

Query Match 9.0%; Score 69.5; DB 1; Length 474;
Best Local Similarity 28.6%; Pred. No. 34;
Matches 42; Conservative 16; Mismatches 58; Indels 31; Gaps 8;

QY 19 LCKALSQSLPYLYRPOSSLSFLITDIKKIDIOYFLPLTGGKCLHLRLTGQAFCLILEF 78

Db 215 LSKAIASLEEI---AQRLSKVLA--TALDLYGSLNVAHCLQ-RLRQYGDVY-- 266

QY 79 LPWCNGIIESLVGLENER-----KVLSSGSSQQM---GEARFGMEWEVFPLELG 124

Db 267 FSWGNGQDCFGASPERLLSLHNLQVTDALAGSAPRDVDQGRDLGQGLLHNPXELR 326

QY 125 RPEA---GALQRL-----PQNSALL 142

Db 327 EHQAVLDYLLQRLRALGLSPQASSLKL 353

RESULT 13

H64476

lipote protein ligase (EC 3.4.21.-) *Lon* - *Methanococcus jannaschii*

N;Alternate names: Alp-dependent proteinase *lon* homolog

C;Species: *Methanococcus jannaschii*

C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004

C;Accession: H64476

R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak

ison, J.D.; Sadov, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,

A;Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*

A;Reference number: A64300; MUID:96337999; PMID:8688087

A;Accession: H64476

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-649 <BUL>

C;Genetics:

A;Map position: FOR1378485-1380434

C;Superfamily: Lon-type protease

C;Keywords: ATP; hydrolase; nucleotide binding; P-loop; serine proteinase

F;6-89/Domain: *Methanococcus* endopeptidase La homolog P-loop-containing homology <MLA

F;47-54/Region: nucleotide-binding motif A (P-loop)

F;238-243/Region: nucleotide-binding motif B

F;550/Active site: Ser #status predicted

Query Match 9.0%; Score 69.5; DB 1; Length 649;
Best Local Similarity 27.6%; Pred. No. 47;
Matches 42; Conservative 24; Mismatches 53; Indels 33; Gaps 10;

QY 2 VVFTLLVLL--LLLLLTALCKALSQSLPYLYRPOSSLSFLITDIKKIDIOYFLPLT 59

Db 143 VTITLVIVFIFGVIIITLSIMGASRSMNNL-NPM-DLKPVLVECK-----RLPVR 194

QY 60 GKCLHL-RLTGQRAFCILFPLPWCNGIIESLVGLENERKVLSSGSSQQMGEGARM---- 114

Db 195 ASAYNTRLLGDIKHCPGLGRP-----PLGTPPKRIILGA----IHEAHRGILYVD 242

QY 115 EWEVFFLEGRPEAGALQ--RLP-----QPNSA 140

Db 243 EIKTMPELVQDYILTALQDKQLPISGRNPNS 274

RESULT 14

S70786

probable transcription regulator *csgD* - *Escherichia coli* (strain K-12)

C;Species: *Escherichia coli*

C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 09-Jul-2004

C;Accession: S70786; E64846

R;Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A;Title: Expression of two *csg* operons is required for production of fibronectin- and

A;Reference number: S70783; MUID:96414468; PMID:8817489

A;Accession: S70786

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-216 <HAM>

A;Cross-references: UNIPROT:P52106; UNIPARC:UPI000012852F; EMBL:X90754; NID:gl147558;

A;Experimental source: strain K12, substrain MC4100

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of *Escherichia coli* K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: E64846

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-216 <BLAT>

A;Cross-references: UNIPARC:UPI000012852F; GB:AE000205; GB:U00096; NID:gl787265; PIDN

A;Experimental source: strain K-12, substrain MG1655

Query Match	9.0%;	Score 69;	DB 2;	Length 253;
Best Local Similarity	23.0%;	Pred. No. 20;		
Matches 35;	Conservative 32;	Mismatches 55;	Indels 30;	Gaps 7;
Qy	7	LLVLLLLLLLTAACKALQSQSLPVTLYXRPQSSFLLTIDIKKIDIQYFLPLT-----	58	
Db	113	LLVYLSIQI VALK-----LDPSLY-----TIFLTSSNVLSWSLSYLII LSTISPIPESN	163	
Qy	59	GGKCLHLRLTGORAFCI LEPWCN--GIATESVLGLENERKVL S--GGSSQMGAEARRM	114	
Db	164	LGRIFHDKLAKGKSTILLLELIPISLYNVVIFIFYIMYKICGSLGGSSGDIAGASITL	223	
Qy	115	EWEVFPELGRPEAGALQRLPQN SALLACRC	146	
Db	224	S---FPFLITDEITNL-----NYSLSILC	246	

Search completed: January 10, 2007, 15:06:43
Job time : 40 secs


```

; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2600
; LENGTH: 415
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2600

```

```

Query Match      9.8%; Score 75.5; DB 2; Length 415;
Best Local Similarity 24.6%; Pred. No. 1.9;
Matches 32; Conservative 23; Mismatches 44; Indels 31; Gaps 5;

QY    5 TLLVLLLLLALCKALSPLTYLRPOSSLSF-----LITDIKKIDIOFYPLT 58
       :|::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db    43 TVLTSLSLIICISLLMIASIPFSAAADLASRFFWQLAYVVGTMAMIVRP-- 100
       :|::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

QY    59 GGKCLHLRTGQAFCTILEFLPCWG----IIESLVGLENERKVLGGSSQQMGARRGM 114
       :|::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db   101 -----LRVYYQRTHINVFVLMVMVALGLLILTAMFG-----DING-----SRRWL 141
       :|::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

QY   115 EWEVFPLELG 124
       :|::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db   142 DLGIENLOAG 151

```

```

RESULT 3
US-09-949-016-8503
; Sequence 8503, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWLEDGE OF
; HUMAN DISEASE, WITH HUMAN DISEASE, P
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIORITY APPLICATION NUMBER: 60/241,755
; PRIORITY FILING DATE: 2000-10-20
; PRIORITY APPLICATION NUMBER: 60/237,768
; PRIORITY FILING DATE: 2000-10-03
; PRIORITY APPLICATION NUMBER: 60/231,498
; PRIORITY FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 8503
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-8503

```

Query Match	9.5%	Score 73.5;	DB 2;	Length 256;
Best Local Similarity	22.9%	Pred. No. 1.7;		
Matches	32;	Conservative 12;	Mismatches 45;	Indels 51; Gaps 4;
Qy	12	LLLLLTALCKALSOSLYTLRPOSSLSFLIITDIKKIDIOYFLPTGKGKHLHLRLTQSR	71	
Db	13	LLLLYIIIEPDYLGNNHPTCKPOENMG-----RVFLTGEK	48	
Qy	72	AFCLTEFLPCWNCIIIESL-----VGLENERKV---LSGSSSQMG	108	
Db	49	ANSILKRYPRANGFFEEIRQGNIERECKEEFCTFEAREAFENNEKTFEWSYTKAQOG	108	
Qy	109	EARGMEW-----EVFPLELG	124	
Db	109	ESNREGSDWFOFLYTFPLIFG	128	

RESULT 4
US-09-675-305-6
; Sequence 6, Application US/09675305
; Patent No. 641151
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander J

```

; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur J.
; TITLE OF INVENTION: SEQ. 6441153el Human Carboxypeptidases and
; FILE OF INVENTION: POLYNUCLEOTIDES Encoding the Same
; FILE REFERENCE: LEX-0047-USA
; CURRENT APPLICATION NUMBER: US/09/675,305
; CURRENT FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: US 60/156,685
; PRIOR FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 247
; TYPE: PRT
; ORGANISM: homo sapiens
; US-09-675-305-6

```

[illegible]

```

RESULT 5
US-10-200-344-6
; Sequence 6, Application US/10200344
; Patent No. 6780640
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6780640el Human Carboxypeptidases and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0047-USA
; CURRENT APPLICATION NUMBER: US/10/200.344
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: US/09/675,305
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 247
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-200-344-6

```

```

Query Match      9.5%; Score 73; DB 2; Length 247;
Best Local Similarity 29.5%; Pred. No. 1.9;
Matches 31; Conservative 16; Mismatches 30; Indels 28; Gaps 6

Qy      21 KALSOSPYLYRPOSSLFL---LITDIKKIDIOYELPLTTGGKCLHLRLITGQAFICILE 77
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      57 KKISYQLKVLWQP-SSISYVSEGTVDV-----HIPQNGSRA--LLA 96

Qy      78 FLPCWNGIIESLVGLENERKVLGGSSQQMGEAR---GMEWEVVF 119
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      97 FLOEAN--IQYKVLIEDLQTKEGSSLTQRNRRLSGYNYEYV 139

RESULT 6
US-09-675-305-12

```

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	770	100.0	150	5	AAU83189	Aau83189 Novel sec
2	770	100.0	150	5	AAW50708	Aam50708 Human sec
3	164.5	21.4	246	4	ABG29473	Abg29473 Novel hum
4	134	17.4	635	4	ABG19492	Abg19492 Novel hum
5	134	17.4	2134	4	ABG27624	Abg27624 Novel hum
6	132.5	17.2	138	4	ABG50801	Abg50801 Novel hum
7	124.5	16.2	313	4	ABB12469	Abb12469 Human bon
8	96	12.5	216	8	ADU02421	Adu02421 Novel hum
9	95.5	12.4	87	4	ABG23765	Abg23765 Novel hum
10	93	12.1	87	4	ABG12819	Abg12819 Novel hum
11	84	10.9	263	4	ABG24127	Abg24127 Novel hum
12	84	10.9	263	4	ABG07958	Abg07958 Novel hum
13	79.5	10.3	459	8	ADP30331	Adp30331 Human sec
14	77.5	10.1	322	7	ABG36760	Abg36760 Rhesus ro
15	75.5	9.8	391	6	ABU35437	Abu35437 Protein ec
16	75.5	9.8	415	8	ADL04914	Adl04914 M. catarr
17	75	9.7	527	8	ADN26981	Adn26981 Bacterial
18	74	9.6	137	4	AAU48296	Aau48296 Propionib
19	74	9.6	137	6	ABM44815	Abm44815 Propionib
20	74	9.6	307	8	ADH42051	Adh42051 Novel hum
21	74	9.6	1105	7	ABM50228	Abm50228 Rice abio
22	74	9.6	1218	7	ABM88074	Abm88074 Rice abio
23	73.5	9.5	414	7	ADP03569	Adp03569 Human gpc

CC reagents for the study of cells, receptors, and other binding molecules.
CC The polynucleotide is useful for radiation hybrid mapping, and somatic
CC cell genetic technique developed for constructing high-resolution,
CC contiguous maps of mammalian chromosomes. Reagents disclosed in the
CC invention may be used to detect metabolic abnormalities characterised by
CC over or under production of the protein. This is the amino acid sequence
CC of a mammalian secreted polypeptide, described in the method of the
CC invention
XX
SQ Sequence 150 AA;
Query Match 100.0%; Score 770; DB 5; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.8e-78;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MVVFTLLVLLVLLLTALCKALQSILPYTYLRPOSSLSFLITDIKKIDIQYFLPTGG 60
Db 1 MVVFTLLVLLVLLLTALCKALQSILPYTYLRPOSSLSFLITDIKKIDIQYFLPTGG 60
Qy 61 KCLHLRLTGQAFCLIFLPWCNGIIESLVGLNERKVLSCGSSQOQGEARRGMEWEVFP 120
Db 61 KCLHLRLTGQAFCLIFLPWCNGIIESLVGLNERKVLSCGSSQOQGEARRGMEWEVFP 120
Qy 121 LELGRPEAGALQRLPQPNLSALLACRCAGY 150
Db 121 LELGRPEAGALQRLPQPNLSALLACRCAGY 150
RESULT 2
AM50708
ID AM50708 standard; protein; 150 AA.
AC
XX
XX
XX 08-APR-2002 (first entry)
XX Human secreted cytokine Zcyto27.
XX
XX Zcyto27; MSP; secreted protein; cytokine; human; antiinflammatory;
XX immunomodulator; therapy.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..22
XX Region /label= Signal_Peptide
XX Protein 22..35
XX /label= Helix-A
XX 23..150
XX /label= Mature_Protein
XX /note= "the mature protein is separately claimed in Claim
XX 1"
XX Region 60..73
XX /label= Helix-B
XX Region 84..95
XX /label= Helix-C
XX Region 113..130
XX /label= Helix-D
XX
XX WO200202626-A2.
XX
XX 10-JAN-2002.
XX
XX 29-JUN-2001; 2001WO-US020847.
XX
XX 30-JUN-2000; 2000US-0215446P.
XX
XX (ZYMO) ZYMOGENETICS INC.
XX
XX Sheppard PO, Presnell SR, Taft DW;
XX
XX WPI; 2002-154728/20.
XX
XX N-PSDB; ABA91403, ABA91404.
XX

XX New polypeptides which include cytokines, growth factors and secreted
XX proteins, useful for modulating immune reaction and inflammation and
XX preventing or treating pancreatic disorders, diabetes and degenerative
XX diseases.
XX
XX Claim 2; Page 55-56; 61pp; English.
XX
XX The present sequence is that of Zcyto27, a novel human cytokine. Zcyto27,
XX the mature portion of which is also claimed, is a secreted protein and a
XX member of a novel group of proteins termed MSP. It bears significant
XX identity to interleukin-2 (12%), interleukin-3 (13%) and interleukin-5
XX (12%). These cytokines mediate immunological responses, including
XX inflammation. As a protein that is expressed in kidney, liver and foetal
XX liver, peripheral blood lymphocytes, prostate, spleen, testis, thymus,
XX lymph node and bone marrow, Zcyto27 polypeptides and polynucleotides will
XX find use in modulating the immune reaction and inflammation in these and
XX other organs. Recombinant Zcyto27 polypeptides can be produced in host
XX cells. They can be used to raise specific antibodies, and to screen for
XX agonists and antagonists. They may also be produced as part of a fusion
XX protein e.g. with a toxin for targeted cell or tissue ablation, and
XX cancer therapy. The signal peptide can be used to direct a fused protein
XX through the secretory pathway of a cell
XX
XX Sequence 150 AA;
Query Match 100.0%; Score 770; DB 5; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.8e-78;
Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MVVFTLLVLLVLLLTALCKALQSILPYTYLRPOSSLSFLITDIKKIDIQYFLPTGG 60
Db 1 MVVFTLLVLLVLLLTALCKALQSILPYTYLRPOSSLSFLITDIKKIDIQYFLPTGG 60
Qy 61 KCLHLRLTGQAFCLIFLPWCNGIIESLVGLNERKVLSCGSSQOQGEARRGMEWEVFP 120
Db 61 KCLHLRLTGQAFCLIFLPWCNGIIESLVGLNERKVLSCGSSQOQGEARRGMEWEVFP 120
Qy 121 LELGRPEAGALQRLPQPNLSALLACRCAGY 150
Db 121 LELGRPEAGALQRLPQPNLSALLACRCAGY 150
RESULT 3
ABG29473
ID ABG29473 standard; protein; 246 AA.
AC
XX
XX
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #29464.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS93660.
XX

XX	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity.
XX	
PS	Claim 20; SEQ ID NO 59832; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC	sequences. (I) is useful as hybridisation probes, polymerase chain
CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC	and in recombinant production of (II). The polynucleotides are also used
CC	in diagnostics as expressed sequence tags for identifying expressed
CC	genes. (I) is useful in gene therapy techniques to restore normal
CC	activity of (II) or to treat disease states involving (II). (II) is
CC	useful for generating antibodies against it, detecting or quantitating a
CC	polypeptide in tissue, as molecular weight markers and as a food
CC	supplement. (II) and its binding partners are useful in medical imaging
CC	of sites expressing (II). (I) and (II) are useful for treating disorders
CC	polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC	amino acid sequences of the invention. Note: The sequence data for this
CC	patent did not appear in the printed specification, but was obtained in
CC	electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 246 AA;
	Query Match 21.4%; Score 164.5; DB 4; Length 246;
	Best Local Similarity 60.3%; Pred. No. 9.7e-10;
	Matches 47; Conservative 2; Mismatches 22; Indels 7; Gaps 4;
OY	64 HURLTCORAFCLFPLWNCNGIIESLVGLENERKV-LSGSSQQWGEARRGMWEVFPPL 121
DB	58 HSLSQGORAFCILGLFPHSRIRSHVGLNECKVLLSGRSSQQMGEPXGRFXSPXVGFL 117
OY	122 ELGRPAGALQRLPQNS 139
DB	118 G-GR----ALLQLPQNS 130
RESULT 4	
ABG19492	ID ABG19492 standard; protein; 635 AA.
AC	ABG19492;
DT	13-FEB-2002 (first entry)
DE	Novel human diagnostic protein #19483.
XX	Human; Chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
OS	Homo sapiens.
XX	WO200175067-A2.
PN	11-OCT-2001.
PD	
XX	30-MAR-2001; 2001WO-US008631.
PF	
XX	31-MAR-2000; 2000US-00540217.
PR	23-AUG-2000; 2000US-00649167.
XX	(HYSE-) HYSEQ INC.
PA	
XX	Dzmanac RT, Liu C, Tang YT;
P1	
XX	

DR	WPI; 2001-639362/73.
DR	N-PSDB; AAS83679.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity.
XX	
PS	Claim 20; SEQ ID NO 49851; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC	sequences. (I) is useful as hybridisation probes, polymerase chain
CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC	and in recombinant production of (II). The polynucleotides are also used
CC	in diagnostics as expressed sequence tags for identifying expressed
CC	genes. (I) is useful in gene therapy techniques to restore normal
CC	activity of (II) or to treat disease states involving (II). (II) is
CC	useful for generating antibodies against it, detecting or quantitating a
CC	polypeptide in tissue, as molecular weight markers and as a food
CC	supplement. (II) and its binding partners are useful in medical imaging
CC	of sites expressing (II). (I) and (II) are useful for treating disorders
CC	polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC	amino acid sequences of the invention. Note: The sequence data for this
CC	patent did not appear in the printed specification, but was obtained in
CC	electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 635 AA;
	Query Match 17.4%; Score 134; DB 4; Length 635;
	Best Local Similarity 48.1%; Pred. No. 8.9e-06;
	Matches 38; Conservative 8; Mismatches 27; Indels 6; Gaps 3
OY	76 LEFLPWNCNGIIESLVGLENERKV-LSGSSQQWGEARRGMWEVFPLELGRPEAGALQRL 134
DB	399 LGFLPWCTRRIHSHLGLENECVLLSGSSSQMKPEGRWSGVKVP-GVGPLSSWTLLQL 134
OY	135 PQPNMAL---LACRCAGA 149
DB	458 PWPNSALFCRWMACLSAGS 476
RESULT 5	
ABG27624	ID ABG27624 standard; protein; 2134 AA.
AC	ABG27624;
DT	18-FEB-2002 (first entry)
DE	Novel human diagnostic protein #27615.
XX	Human; Chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
OS	Homo sapiens.
XX	WO200175067-A2.
PN	11-OCT-2001.
PD	
XX	30-MAR-2001; 2001WO-US008631.
PF	
XX	31-MAR-2000; 2000US-00540217.
PR	23-AUG-2000; 2000US-00649167.
XX	(HYSE-) HYSEQ INC.
PA	
XX	
P1	
XX	

PI Dmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS91811.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 57983; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activities. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2134 AA;
 Query Match 17.4%; Score 134; DB 4; Length 2134;
 Best Local Similarity 48.1%; Pred. No. 4.2e-05;
 Matches 38; Conservative 8; Mismatches 27; Indels 6; Gaps 3;
 QY 76 LEFLPWCNGIIESLVGLENERKV-LSGGSSQQMGEGARRGMWVFPFLGLRPEAGALQRL 134
 DB 881 LGFLPWCNTRIGSHVLENECVLLSGSSQQMGKEPEGRMSGKVSP-GVGPLSSWTLQL 939
 QY 135 POPNSAL----LACRCAGA 149
 DB 940 FWPNSALFCRWMACLSAGS 958
 RESULT 6
 ID ABG05801 standard; protein; 138 AA.
 XX
 AC ABG05801;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #5792.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX

(HYSE-) HYSEQ INC.
 XX
 PI Dmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS69988.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 36160; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activities. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 138 AA;
 Query Match 17.2%; Score 132.5; DB 4; Length 138;
 Best Local Similarity 50.6%; Pred. No. 1.9e-06;
 Matches 40; Conservative 1; Mismatches 29; Indels 9; Gaps 4;
 QY 78 FLPCNGIIESLVGLENERKV-LSGGSSQQMGEGARRGMWVFPFLGLRPEAGALQRLP 135
 DB 28 FLPCNTRIGSHVLENECVLLSGSSQQMGEGARRGMWVFPFLGLRPEAGALQRLP 87
 QY 136 --QPNSALL-----ACRCA 147
 DB 88 VQTPRCSTCRWPAGACRCA 106
 RESULT 7
 ID ABB12469 standard; protein; 313 AA.
 XX
 AC ABB12469;
 XX
 DT 15-JAN-2002 (first entry)
 XX
 DE Human bone marrow expressed protein SEQ ID NO: 308.
 XX
 KW Human; bone marrow; cytostatic; antirheumatic; antiarthritic; vulnary;
 KW antiinflammatory; antibacterial; immunosuppressive; vasotropic; cancer;
 KW antiparkinsonian; neuroprotective; nootropic; haemostatic; osteopathic;
 KW antitumor; fungicide; antidiabetic; antischismatic; antiallergic;
 KW immunostimulant; analgesic; cerebroprotective; antianemic; infection;
 KW nervous system disorder; autoimmune disorder; inflammation; allergy.
 XX
 OS Homo sapiens.
 XX
 PN WO200174836-A1.
 XX
 PD 11-OCT-2001.

```

XX PF 30-MAR-2001; 2001WO-US010472.
XX PR
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PR 30-NOV-2000; 2000US-0250583P.
XX PA (HYSE-) HYSEQ INC.
XX PT Tang YT, Liu C, Drmanac RT, Ford JE, Boyle BJ;
XX DR WPI; 2001-626375/72.
XX PR
XX PT New bone marrow-expressed nucleic acids and polypeptides, useful for
XX PT diagnosis, treatment of inflammatory, autoimmune, neurological, cancer
XX PT and increasing hematopoiesis, stem cell survival and bone growth and
XX PT remodeling.
XX PS Claim 10; Page 359-360; 380pp; English.
XX CC The present invention relates to bone marrow expressed polynucleotides
XX CC and proteins. These sequences can be used in the treatment of
XX CC inflammatory conditions (eg arthritis, Crohn's disease), cancer, central
XX CC and peripheral nervous system diseases and neuropathies, such as
XX CC Alzheimer's, Parkinson's and Huntington's diseases, spinal cord
XX CC disorders, head trauma, cerebrovascular diseases, myeloid and lymphoid
XX CC cell disorders, platelet disorders, stem cell disorders, bone
XX CC degenerative disorders, autoimmune disorders, for example multiple
XX CC sclerosis, diabetes and arthritis, viral and bacterial infections,
XX CC allergies and blood coagulation disorders. The present sequence is a
XX CC protein of the invention
XX SQ Sequence 313 AA;

Query Match 16.2%; Score 124.5; DB 4; Length 313;
Best Local Similarity 45.9%; Pred. NO. 4.2e-05;
Matches 39; Conservative 3; Mismatches 28; Indels 15; Gaps 5;

QY 60 GKCLHLRLTGQAFCLFELPWCNGHIESLVGLNERKV-LSGSSQQMGARRGMEWEV 118
Db 228 GTWLRSRRTW--SFLYPGGLPWCGRIGRVLGNECKVSLSGSSQPPGEG--RW-- 281

QY 119 FPLELGRPEAGALQRLPQNSALLA 143
Db 282 -----SSPEVGP---LASFGSPLIA 298

RESULT 8
ID ADU02421
XX ADU02421 standard; protein; 216 AA.
XX AC ADU02421;
XX DT 27-JAN-2005 (first entry)
XX DE Novel human polypeptide seqid 888.
XX KW cytostatic; antipsoriatic; antiinflammatory; gene therapy; Nanodisc;
XX KW proliferative disorder; inflammatory disorder; immune disorder;
XX KW metabolic disorder; bone disorder; CNS disorder; cancer; psoriasis;
XX KW ulcerative colitis; human.
XX OS Homo sapiens.
XX PN WO2004093804-A2.
XX PD 04-NOV-2004.
XX PF 19-APR-2004; 2004WO-US012047.
XX PR 18-APR-2003; 2003US-0463708P.
XX PR 18-APR-2003; 2003US-0463732P.
XX PR 02-MAY-2003; 2003US-0467199P.

```

02-MAY-2003; 2003US-0467230P.
19-MAY-2003; 2003US-0471306P.
19-MAY-2003; 2003US-0471336P.
08-JUL-2003; 2003US-0485223P.
08-JUL-2003; 2003US-0485224P.
14-JUL-2003; 2003US-0486446P.
14-JUL-2003; 2003US-0486480P.
08-AUG-2003; 2003US-0493573P.
08-AUG-2003; 2003US-0493577P.
08-SEP-2003; 2003US-0505059P.
(FIVE-) FIVE PRIME THERAPEUTICS INC.
Lee E, Hestir K, Chu K, Masuoka L, Williams LT;
WPI; 2004-775861/76.
N-PSDB; ADU01689.
New first nucleic acid molecule comprising a polynucleotide sequence
given in the specification, useful in preparing a composition for
diagnosing or treating e.g., cancer, psoriasis or ulcerative colitis.
Claim 14; SEQ ID NO 888; 291pp; English.
The invention describes a new first nucleic acid molecule comprising a
polynucleotide sequence given in the specification. Also described are:
an animal injected with the nucleic acid molecule; a second nucleic acid
molecule comprising a second polynucleotide sequence that is at least
about 70, 80, 90 or 95% homologous to the first nucleic acid molecule or
that hybridizes to the first polynucleotide sequence under high
stringency conditions; a vector comprising the nucleic acid molecule and
a promoter that drives the expression of the nucleic acid molecule; a
host cell transformed, transfected, transduced or infected with the
nucleic acid molecule; a nucleic acid composition comprising a carrier or
a buffer and one or more compositions comprising the nucleic acid
molecule, vector or host cell; a substantially purified polypeptide; an
animal injected with the polypeptide; a polypeptide composition
comprising the polypeptide molecule and a carrier or buffer; a cell
culture medium comprising the polypeptide or transfected cells
transfected with the polynucleotide; making a transformed, transfected,
transduced, or infected host cell; synthesizing Nanodiscs simultaneously
and for synthesising a series of simultaneously-synthesised Nanodiscs
sequentially utilising a dynamic system; preparing a hydrophobic protein
for determination of crystal structure; immunising a non-human animal;
screening for modulators of hydrophobic protein activity; a diagnostic
kit; determining the presence of the nucleic acid molecule or its
complement; determining the presence of an antibody to the polypeptide in
a sample; an antibody specifically recognising, binding to or modulating
the biological activity of at least one polypeptide encoded by a nucleic
acid molecule or its biologically active fragment; an antibody
composition comprising the antibody and a carrier; a bacteriophage, where
the antibody is displayed on the bacteriophage; a bacterial cell
comprising the bacteriophage; a non-human animal injected with the
antibody composition; a host cell that secretes the antibody; making an
antibody; diagnosing a disease, disorder, syndrome, or condition
comprising cancer, or proliferative, inflammatory, immune, metabolic,
bone, CNS, genetic, bacterial and viral diseases, disorders, syndromes or
conditions in a patient; a modulator composition comprising a modulator
and a carrier; gene therapy; prophylactic or therapeutic treatment of a
subject; an isolated modified cell comprising at least one first
heterologous nucleic acid molecule, where the first heterologous nucleic
acid molecule comprises a first polynucleotide sequence that encodes a
first polypeptide; a non-human animal deficient in the polypeptide or
that over-expresses the polypeptide; isolated tissues derived from the
non-human animal; and one or more cells derived from the non-human
animal. The nucleic acid is useful in preparing a composition for
diagnosing or treating e.g., cancer, psoriasis or ulcerative colitis.
This is the amino acid sequence of a novel human polypeptide of the
invention.

Query Match 12.5%; Score 96; DB 8; Length 216;
Sequence 216 AA;
SQ

Best Local Similarity 59.0%; Pred. No. 0.043;		1;	
Matches 23; Conservative 2; Mismatches 10; Indels 4; Gaps		1;	
OY	114 MEWEVFLELGRPEAGALQRLPQNSA 148		
Db	1 MEWEGSPLESGRSAAALFLLLPNSALFROSMACRHAG 39		
RESULT 9			
ID	ABG23765		
XX	ABG23765 standard; protein; 87 AA.		
AC	ABG23765;		
XX	18-FEB-2002 (first entry)		
DT	Novel human diagnostic protein #23756.		
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
DE	food supplement; medical imaging; diagnostic; genetic disorder.		
XX	Homo sapiens.		
XX	WO200175067-A2.		
XX	11-OCT-2001.		
XX	30-MAR-2001; 2001WO-US008631.		
XX	31-MAR-2000; 2000US-00540217.		
XX	23-AUG-2000; 2000US-00649167.		
XX	(HYSE-) HYSEQ INC.		
PI	Drmanac RT, Liu C, Tang YT;		
XX	WPI; 2001-639362/73.		
XX	N-PSDB; AAS87952.		
XX	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity.		
XX	Claim 20; SEQ ID NO 54124; 103pp; English.		
XX	The invention relates to isolated polynucleotide (I) and polypeptide (II)		
CC	sequences. (I) is useful as hybridisation probes, polymerase chain		
CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,		
CC	and in recombinant production of (II). The polynucleotides are also used		
CC	in diagnostics as expressed sequence tags for identifying expressed		
CC	genes. (I) is useful in gene therapy techniques to restore normal		
CC	activity of (II) or to treat disease states involving (II). (II) is		
CC	useful for generating antibodies against it, detecting or quantitating a		
CC	polypeptide in tissue, as molecular weight markers and as a food		
CC	supplement. (II) and its binding partners are useful in medical imaging		
CC	of sites expressing (II). (I) and (II) are useful for treating disorders		
CC	involving aberrant protein expression or biological activity. The		
CC	polypeptide and polynucleotide sequences have applications in		
CC	diagnostics, forensics, gene mapping, identification of mutations		
CC	responsible for genetic disorders or other traits to assess biodiversity		
CC	and to produce other types of data and products dependent on DNA and		
CC	amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic		
CC	amino acid sequences of the invention. Note: The sequence data for this		
CC	patent did not appear in the printed specification, but was obtained in		
CC	electronic format directly from WIPO at		
CC	ftp.wipo.int/pub/published_pct_sequences		
XX	Sequence 87 AA;		
SQ	Query Match 12.4%; Score 95.5; DB 4; Length 87;		
	Best Local Similarity 62.5%; Pred. No. 0.015;		
	Matches 20; Conservative 2; Mismatches 9; Indels 1; Gaps 1;		
Query Match 12.1%; Score 93; DB 4; Length 87;			
Best Local Similarity 43.1%; Pred. No. 0.029;			
Matches 28; Conservative 5; Mismatches 16; Indels 16; Gaps 3;			
OY	90 VGLENERKV-LSGGSSQMGEGARGMEWVFLELGR-PEAGALQRLPQNSALLACR 145		

GenCore version 5.1.9
Copyright (c) 1993 - 2007 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2007, 14:57:46 ; Search time 299 Seconds
(without alignments)
464.055 Million cell updates/sec

Title: US-10-777-578-2
Perfect score: 770
Sequence: 1 MVVFTLLVLLLLLLTALC.....LQRLPQNSALLACRCAGAY 150

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_7.2.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87.5	11.4	483	Q4HI99_CAMCO	Q4hi99 campylobact
2	85.5	11.1	550	Q8XR68_RALSO	Q8xr68 ralstonia s
3	83.5	10.8	254	Q97J11_CLOAB	Q97j11 clostridium
4	82	10.6	404	Q9SELI_SOYEN	Q9seli glycine max
5	81	10.5	462	Q743Y7_MYCPA	Q743y7 mycobacteri
6	80.5	10.5	398	Q6AQH5_DESPS	Q6agh5 desulfotale
7	79.5	10.3	470	Q2NAE4_9SPHN	Q2nae4 erythrobact
8	79.5	10.3	953	Q82V63_NITEU	Q82v63 nitrosomona
9	79	10.3	414	Q41U74_FERAC	Q41u74 ferropasma
10	78.5	10.2	673	Q8TFG0_SCHPO	Q8tfgo schizosacch
11	78	10.1	326	Q4VQ72_9REOV	Q4vq72 porcine rot
12	78	10.1	337	Q471X8_RALEJ	Q471x8 ralstonia e
13	78	10.1	392	Q894Q0_CLOTE	Q894q0 clostridium
14	77.5	10.1	105	Q8W2T3_ORYSA	Q8w2t3 oryza sativ
15	77.5	10.1	481	Q4WKL4_ASPPU	Q4wkl4 aspergillus
16	77	10.0	2005	Q5STG5_HUMAN	Q5stg5 homo sapien
17	76.5	9.9	956	Q5PB30_ANAMM	Q5pb30 anaplasmam
18	76.5	9.9	2604	Q50T90_ENTHI	Q50t90 entamoeba h
19	76.5	9.9	5120	PCLO_CHICK	Q9pu36 gallus gall
20	76	9.9	359	Q4J7S5_SULAC	Q4j7s5 sulfolobus
21	76	9.9	431	TOLB_MYXXA	Q84ff7 myxococcus
22	76	9.9	506	Q3QTN9_9RHOB	Q3qtn9 silicibacte
23	76	9.9	604	Q55PQ5_CRYNE	Q55pq5 cryptococcu
24	76	9.9	604	Q5KQD3_CRYNE	Q5kdq3 cryptococcu
25	76	9.9	1086	Q3PWT5_NITHA	Q3pwt5 nitrobacter
26	76	9.9	1097	Q9H6Y0_HUMAN	Q9h6y0 homo sapien
27	75.5	9.8	507	CRNA_EMENI	P22152 emericella
28	75.5	9.8	507	Q5BEM2_EMENI	Q5bem2 aspergillus
29	75.5	9.8	550	Q976S2_SULTO	Q976s2 sulfolobus
30	75	9.7	116	Q5EY93_9EMBE	Q5ey93 ammodramus
31	75	9.7	400	Q9P8B4_AGABI	Q9p8b4 agaricus bi

32	75	9.7	405	2	Q4FNZ4_PELUB	Q4fnz4 pelagibacte
33	75	9.7	497	2	Q5R5P1_PONPY	Q5r5p1 pongo pygma
34	75	9.7	583	2	Q41VE2_DESHA	Q41ve2 desulfitoba
35	74.5	9.7	319	2	Q4B7L2_BURVI	Q4b7l2 burkholderi
36	74.5	9.7	423	2	Q32P40_HUMAN	Q32p40 homo sapien
37	74.5	9.7	1153	2	Q7XUT7_ORYSA	Q7xut7 oryza sativ
38	74	9.6	385	2	Q335T1_CROPO	Q335t1 crocodylus
39	74	9.6	681	2	Q3TTV4_MOUSE	Q3ttv4 mus musculus
40	74	9.6	1159	2	Q62120_ORYSA	Q62120 oryza sativ
41	74	9.6	1188	2	Q69QL2_ORYSA	Q69ql2 oryza sativ
42	74	9.6	1281	2	Q6TJY0_ORYSA	Q6tjy0 oryza sativ
43	74	9.6	1302	2	Q69QM8_ORYSA	Q69qm8 oryza sativ
44	74	9.6	1329	2	Q7RQ32_PLAYO	Q7rq32 plasmodium
45	74	9.6	1479	2	Q3U306_MOUSE	Q3u306 mus musculu

ALIGNMENTS

RESULT 1

Q4HI99_CAMCO PRELIMINARY; PRT; 483 AA.
AC Q4HI99;
DT 16-AUG-2005, integrated into UniProtKB/TrEMBL.
DT 16-AUG-2005, sequence version 1.
DT 07-FEB-2006, entry version 2.
DE Integral membrane protein MviN.
DE Name=mviN; ORFNames=CCO0860;
GN Campylobacter coli RM2228.
OS Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=306254;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=RM2228;
RA Fouts D.E., Mongodin E.F., Mandrell R.E., Miller W.G., Rasko D.A.,
RA Jacques R.J., Brinkac L.M., DeBoy R.T., Parker C.T., Daugherty S.C.,
RA Dodson R.J., Durkin A.S., Madupu R.R., Sullivan S.A., Shetty J.U.,
RA Ayodeji M.A., Shvartsbeyn A.A., Schatz M.C., Badger J.H., Fraser C.M.,
RA Nelson K.E.;
RT "Major structural and novel potential virulence mechanisms from the
RT genomes of multiple Campylobacter species.";
RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.

Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
Distributed under the Creative Commons Attribution-NoDerivs License

CC EMBL; AAFL01000001; EAL57586.1; -: Genomic DNA.
CC GO; GO:0016021; C:integral to membrane, IEA.
CC GO; GO:0009405; P:patogenesis; IEA.
CC InterPro; IPR004268; MVIN_like.
CC Pfam; PF03023; MVIN; 1.
CC PRINTS; PR01806; VIRFACTRMVIN.
CC TIGRPFAM; TIGR01695; mviN; 1.
CC SEQUENCE 483 AA; 54132 MW; 90C4513C4BBE71AB CRC64;

Query Match 11.4%; Score 87.5; DB 2; Length 483;
Best Local Similarity 33.3%; Pred. No. 4.7;

Matches 32; Conservative 13; Mismatches 38; Indels 13; Gaps 4;

QY 5 TLLVLLLLLLTALCKAL-----SQSLPYTLRPSQSL--FLITDIKKIDIQVFLP 56

DB 381 TAAFTAFKALLSALCSLVFLFKDESUKVAVALSSLSAFYLIIANIKERFGKNFFA 440

QY 57 LTGGKCLHLRLTGQAFCL--EFLPWCNGIIESLV 90

DB 441 LISFKFCFLMIGLVVFTLLYEIKPY---LIEALL 473

RESULT 2

OS Q8XR68_RALSO PRELIMINARY; PRT; 550 AA.
 AC Q8XR68;
 DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
 DT 07-FEB-2006, entry version 1.
 DE PROBABLE CYTOCHROME C OXIDASE SUBUNIT TRANSMEMBRANE PROTEIN
 (EC 1.9.3.11).
 DE OrderedLocusNames=RS0993; ORNames=RS02325;
 GN Ralstonia solanacearum (Pseudomonas solanacearum).
 OS Plasmid megaplasmid.
 OC Burkholderiaceae; Ralstonia.
 CC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 CC Burkholderiaceae; Ralstonia.
 CC NCB1_TaxID=305;
 RN [1]
 RT NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 CC STRAIN=GM11000;
 CC MEDLINE=21681879; PubMed=11823852; DOI=10.1038/415497a;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chantier M., Choise N., Claudel-Renard C., Cunac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
 RA Siguier P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weisenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502(2002).
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 CC EMBL; AL646082; CAD18144.1; -; Genomic_DNA.
 DR Biocyc; RSOL305:RSP0993-MONOMER; -;
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0045278; C:respiratory chain complex IV (sensu Bacteria); IEA.
 DR GO; GO:0004129; F:cytochrome-c oxidase activity; IEA.
 DR GO; GO:0016491; F:cytochrome-c oxidase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR004677; COXN.
 DR InterPro; IPR000883; COX1.
 DR PANTHER; PTHR10422:SF1; COXN; 1.
 DR PANTHER; PTHR10422; COX1; 1.
 DR Pfam; PF00115; COX1; 1.
 DR PROSITE; PS0855; COX1; 1.
 DR Complete proteome; Oxidoreductase; Plasmid; Transmembrane.
 KW SEQUENCE 550 AA; 58994 MW; 3D8A20B5FF35C4C CRC64;
 SQ SEQUENCE 550 AA; 58994 MW; 3D8A20B5FF35C4C CRC64;
 Query Match 11.1%; Score 85.5; DB 2; Length 550;
 Best Local Similarity 28.6%; Pred. No. 8.6;
 Matches 48; Conservative 19; Mismatches 62; Indels 39; Gaps 9;
 QY 1 MVVFTLL---LVLLLLLLLTKALQSLSPLTYRPSQSLFLITDKIDIQY-FLP 56
 DB 94 LVVFFVLCAMVWLLVASLAGITASLKLHPDAL-ATQAWLSFGRIITHLNAVAYGAP 152
 QY 57 LTG-----GKCHLHLTGORATCIIEFLPWCNGIIESL-----VGLNEKVLGGSS 104
 DB 153 MAGLGIAQFVLPVRLKPLVGGN-FAVLGAMMNVAGVGLSGIAGVLSL----- 201
 QY 105 QQMGARGMEVPEFLGRPEA--GALQRLPQPNALLACRCAGAY 150
 DB 202 -----GLEWLEIPWQIGVLFAVGGLVGIPLV-LTLLATRVSHLY 240
 RESULT 3
 Q97J11_CLOAB PRELIMINARY; PRT; 254 AA.
 ID Q97J11_CLOAB
 AC Q97J11;
 DT 01-OCT-2001, integrated into UniProtKB/TrEMBL.
 DT 01-OCT-2001, sequence version 1.
 DT 07-FEB-2006, entry version 27.
 DE Proline/glycine betaine ABC transport system, ATPase component.
 GN OrderedLocusNames=CAC1475;
 OS Clostridium acetobutylicum.
 CC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 CC Clostridium.
 CC NCB1_TaxID=1488;
 RN [1]
 RT NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 CC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 CC MEDLINE=21359325; PubMed=11466286;
 CC DOI=10.1128/JB.183.16.4823-4838.2001;
 RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
 RA Tabson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatunov R.L., Sabathe F., Doucette-Stamm L.A., Soucaille P.,
 RA Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 bacterium Clostridium acetobutylicum";
 RL J. Bacteriol. 183:4823-4838(2001).
 CC -----
 CC -1- SIMILARITY: Belongs to the ABC transporter family.
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 CC EMBL; AE007658; AAK79443.1; -; Genomic_DNA.
 DR PIR; H97081; H97081.
 DR HSP; Q9YGA6; IG29.
 DR Biocyc; CAC1488:CAC1475-MONOMER; -;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0016887; F:ATPase activity; IEA.
 DR GO; GO:0000166; F:nucleotide binding; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transporter; IEA.
 DR InterPro; IPR001452; SH3_TRANSPORTER_1.
 DR Pfam; PF00005; ABC_tran; 1.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000006; ABC_transporter; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 KW ATP-binding; Complete proteome; Membrane; Nucleotide-binding;
 KW Transport.
 SQ SEQUENCE 254 AA; 28846 MW; 5172AEB84EEAD3BA CRC64;
 Query Match 10.8%; Score 83.5; DB 2; Length 254;
 Best Local Similarity 32.5%; Pred. No. 6;
 Matches 37; Conservative 10; Mismatches 42; Indels 25; Gaps 6;
 QY 17 TALCKALSQSLPYTYRPSQSLFLITDKIDIQYFLPLTG-----GKCHLHLTGQ 70
 DB 45 TTLKLMNR-----LYEPDKGSIVLFNEDIKKIDVKKLRRSIGVYIQVGLPHTMIANN 99
 QY 71 RAFCILEFLPWCNGIIES-----LVGLE-NERK-----VLSGSSSQMGEAR 111
 DB 100 IA-TVPKLLKWDREIEKRIEDELHLVGLPNEPKKRYPSQLSGGQQQRIGLAR 152
 RESULT 4
 Q9SELI_SOYBN PRELIMINARY; PRT; 404 AA.
 ID Q9SELI_SOYBN
 AC Q9SELI;
 DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.
 DT 01-MAY-2000, sequence version 1.
 DT 07-FEB-2006, entry version 14.
 DE Seed maturation protein PM23 (Fragment).
 GN Name=PM23;
 OS Glycine max (Soybean).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 CC Rosids; eurosids 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 CC Glycine.
 CC NCB1_TaxID=3847;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.

```

RA Hsing Y.I.C., Lin T.Y., Lin T.Y., Liu S.M., Chow T.Y.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF116752; AAF211309.1; -, mRNA.
FT NON TER
SQ SEQUENCE 404 AA; 45221 MW; 684D285911DEF19E CRC64;

Query Match      10.6%; Score 82; DB 2; Length 404;
Best Local Similarity 23.0%; Pred. No. 14;
Matches 41; Conservative 33; Mismatches 44; Indels 60; Gaps 8;

QY 6 LLLVLLLLLLTALCKALQSLEY-----TLRPOSSLSFL----- 41
DB 19 VLVVVLLIGMATLSLSSLPFLHRYNPNPNSLFRPSLSFSFKRSPFLVLAASS 78
QY 42 -----LITDIKKIDIQVF--LPLTGGKCLHLRLTGQAFCLIFLFP--WC 82
DB 79 HDPASNSKSVLTLEIQLEPLDVSHIQKDVPTTADAMKRTISG-----MLGLFSDQF 133
QY 83 NGHIESVLGNERNKVLGGSSQOMGEARRGMW-----EVFLELGRPEAGALQ 132
DB 134 HVVIEAL--WEPLSKLLI--SSMWTGYTLRNVEYRLCLEKNLDMFEGDIEKPKAESMK 187

RESULT 5
Q743Y7 MYCPA
ID Q743Y7 MYCPA PRELIMINARY; PRT; 462 AA.
AC Q743Y7;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 11.
DE Hypothetical protein.
GN OrderedLocusNames=MAP0454; ORFNames=MAP_0454;
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinomycetidae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium avium complex (MAC).
OX NCBI_TaxID=1770;
RN (1)
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RA Li L., Bannantine J., Zhang Q., Amonsin A., Alt D., Kapur V.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AE016958; AAS02771.1; -, Genomic_DNA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 462 AA; 50122 MW; 437D61E14569F7AD CRC64;

Query Match      10.5%; Score 81; DB 2; Length 462;
Best Local Similarity 25.2%; Pred. No. 21;
Matches 38; Conservative 25; Mismatches 52; Indels 36; Gaps 6;

QY 2 VVFTLLVLLLLLALCKALQSPLTYLYRQSSLSFLITDIKKI-DIQVFLPTG 60
DB 36 LVFTNRVELLKLAVILAAVAGAFVSVLYRRQSDAQSRVRLKLVYDLQ----- 87
QY 61 KCLHLRLTGQAFCLIFLWPNCNGIIESLVGLNERKVLGGSSQ-----QMGEA 110
DB 88 --LDREISARREY-----ELTVESQLRELASELRQAQADDLAEIRALSAL 132
QY 111 RRGMEVEVPLEIG-RPEAGALQRLPQNSA 140
DB 133 RTSLE-ILFDTDLGQALQAFGEFQPERA 162

RESULT 6
Q6AQH5_DESPS

```

```

ID Q6AQH5_DESPS PRELIMINARY; PRT; 398 AA.
AC Q6AQH5;
DT 13-SEP-2004, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE Probable Na+/H+ antiporter.
GN OrderedLocusNames=DP0669;
OS Desulfotalea psychrophila.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfobacteriales;
OC Desulfobulbaceae; Desulfotalea.
OX NCBI_TaxID=84980;
RN (1)
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=LSV54 / DSM 12343;
RX PubMed=15305914; DOI=10.1111/j.1462-2920.2004.00665.x;
RA Rabus R., Ruepp A., Frickey T., Rattei T., Fartmann B., Stark M.,
RA Bauer M., Zibat A., Lombardot T., Becker I., Amann J., Gellner K.,
RA Teeling H., Leuschner W.D., Gloeckner F.-O., Lupas A.N., Amann R.,
RA Klenk H.-P.;
RT "The genome of Desulfotalea psychrophila, a sulfate-reducing bacterium
RT from permanently cold Arctic sediments.";
RL Environ. Microbiol. 6:887-902(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; CR522870; CAG35398.1; -, Genomic_DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0006885; P:regulation of pH; IEA.
DR GO; GO:0006814; P:sodium ion transport; IEA.
DR InterPro; IPR004670; NhaA.
DR Pfam; PF06965; Na H antiport 1; 1.
DR TIGRFAMs; TIGR00773; NhaA; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 43064 MW; 7571BD49177DC42E CRC64;

Query Match      10.5%; Score 80.5; DB 2; Length 398;
Best Local Similarity 28.2%; Pred. No. 20;
Matches 42; Conservative 18; Mismatches 42; Indels 47; Gaps 7;

QY 3 VFTLL-----LVLLLLLLTALCKALQSPLTYLYRQSSLSFLITDIKKIDIQVFLPT 58
DB 10 VFALLKSNISGGILLMLATALALIMANSPGHVLY-----SMLITTPVE---VRFGLPE 59
QY 59 GSKCLHLRLTGQAFCLIFLWPNCN-GIIES---LVGLNERKVLGGSSQOMGEARRGM 114
DB 60 TAKPL-----LLWINDGLMAGFFLVGLKREIFEGGLSQRS----- 97
QY 115 EWEVPLELGRPEAGALQRLPQNSALLA 143
DB 98 -----NILLPAIGALGMVVPSCIYLA 119

RESULT 7
Q2NAE4_9SPHN
ID Q2NAE4_9SPHN PRELIMINARY; PRT; 470 AA.
AC Q2NAE4;
DT 07-FEB-2006, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, sequence version 1.
DT 07-FEB-2006, entry version 1.
DE NtrC, nitrogen regulation protein NtrC.
GN ORFNames=ELI_06275;
OS Erythrobacter litoralis HTCC2594.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
OC Erythrobacteraceae; Erythrobacter.
OX NCBI_TaxID=314225;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HTCC2594;
RA Giannoni S.J., Cho J.-C., Ferreira S., Johnson J., Kravits S.,
RA Halpern A., Remington K., Beeson K., Tran B., Rogers Y.-H.,
RA Friedman R., Venter J.C.;
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.

```

```
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000157; ABCG3347.1; -, Genomic DNA.
SQ SEQUENCE 470 AA; 51111 MW; 7DC67C079538405F CRC64;

Query Match 10.3%; Score 79.5; DB 2; Length 470;
Best Local Similarity 27.7%; Pred. No. 30;
Matches 28; Conservative 13; Mismatches 37; Indels 23; Gaps 4;

Qy 52 QYFLPLTGGKCLHLRTGQAFCLFPLPWCNGIIE-----SLVGLNERK 97
Db 320 RHFLALAADEGLPTIDEAIAVLROQWGNVRELVYRLALLARNDVIDALAE 379

Qy 98 VL--SGSSSQOMGEA---RRGMEWEVFPLEGRPEAGALOR 133
Db 380 VLDHEGGPSTMEGPPSVDRAMLDW-----IENRRPAPGALYR 416

RESULT 8
Q82V63 NITEU PRELIMINARY; PRT; 953 AA.
AC O82V63;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 14.
DE Putative oxygenase.
GN OrderedLocusNames=NE1238;
OS Nitrosomonas europaea.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_TaxID=915;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 19718 / IFO 14298;
RX MEDLINE=22586410; PubMed=12700255;
RX DOI=10.1128/JB.185.9.2759-2773.2003;
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.L.,
RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA Arciero D.M., Hommes N.G., Whittaker M.M., Arp D.J.;
RT "Complete genome sequence of the ammonia-oxidizing bacterium and
RT obligate chemolithoautotroph Nitrosomonas europaea.";
RL J. Bacteriol. 185:2759-2773 (2003).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BX321850; CAD85149.1; -, Genomic DNA.
DR Biocyc; NEUR228410:NE1238-MONOMER; -.
GO GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR002007; Aniperoxidase.
DR Pfam; PF03098; An_peroxidase; 1.
DR PROSITE; PS50292; PEROXIDASE_3; 1.
KW Complete proteome.
SQ SEQUENCE 953 AA; 108546 MW; B2461487B7E3EA66 CRC64;

Query Match 10.3%; Score 79.5; DB 2; Length 953;
Best Local Similarity 31.2%; Pred. No. 65;
Matches 44; Conservative 16; Mismatches 38; Indels 43; Gaps 9;

Qy 4 FTLLVLLVLLLLLTLKALQSPLTYLYRPOSSLSFLITDIK---KIDIQYFLPTGG 60
Db 4 FPIVIVLGLLLSG-CDALEPEI-----SCLSVLMQGDIOHVAKTREQRFU-----G 50

Qy 61 KCLHLRLTGQAFCL-----LEFLPCN-----GITESLVGLNERKVLGG--SSQ 106
Db 51 K-----VTGRRACGGDHAVALNRPWLDFNFWGTGDSL-----SLSSSLASSF 97

Qy 107 MGEARRGMEWEVFPLEGRPE 127
Db 98 FGPNERGINSALVELELQRIE 118
```

```
RESULT 9
Q41U74 FERAC PRELIMINARY; PRT; 414 AA.
AC Q41U74;
DT 27-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 27-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Hypothetical protein.
GN ORFNames=FacidRAFT_1367;
OS Ferroplasma acidimanus Ferl.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Ferroplasmaceae; Ferroplasma.
OX NCBI_TaxID=333146;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Ferl;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Ferroplasma
RT acidimanus ferl.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Ferl;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Ferroplasma acidimanus
RT ferl.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Ferl;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAC04000003; EAM94257.1; -, Genomic DNA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001046; Nramp.
DR ProDom; PD001861; Nramp; 1.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 414 AA; 44916 MW; 2A80611B1E404133 CRC64;

Query Match 10.3%; Score 79; DB 2; Length 414;
Best Local Similarity 22.3%; Pred. No. 30;
Matches 33; Conservative 30; Mismatches 57; Indels 28; Gaps 5;

Qy 1 MWFTLLVLLVLLLLLTLKALQSPLTY---LYRPOSSLSFLITDIKIDIQYFL-- 55
Db 153 LIIISLLIIALLLSL-ALRGIIFVSSPLANPILIKTAGYFFLLAANVGAVIMPFMIF 211

Qy 56 -----PLTGKCLHLRTGQAFCLFPLPWCNGIIESLV-----GLENER 96
Db 212 QASATGKLGLDGG---HIRTRRSLRIRKETLAGIVTELLMWIAEMAFAGIPHAS 268

Qy 97 KVLGGSSQQMGGEARRGMEWEVFPLELG 124
Db 269 HSSFFATPQQLGKVLIPVAGDFSPILG 296

RESULT 10
Q8TFG0_SCHPO
```


